DOCUMENT RESUME

ED 058 304 TM 001 019

AUTHOR	Sharon, Amiel T.
TITLE	Test of English as a Foreign Language as a Moderator of Graduate Record Examinations Scores in the Prediction of Foreign Students' Grades in Graduate School.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	RB-71-50
PUB DATE	Sep 71
NOTE	20p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Aptitude Tests; *English (Second Language); *Foreign Students; Grade Point Average; *Grade Prediction; *Graduate Students; Multiple Regression Analysis; *Predictive Validity; Scores; Success Factors; Verbal Ability
IDENTIFIERS	Graduate Record Examinations; *Test of English as a Foreign Language

ABSTRACT

Scores of foreign graduate students on the Graduate Record Examinations (GRE) Aptitude Tests and the Test of English as a Foreign Language (TOEFL) were combined through multiple and moderated regression to predict grade-point average (GPA). It was hypothesized that TOEFL would moderate the relationship between the GRE scores and GPA. According to this hypothesis, students scoring high on TOEFL would be more predictable by GRE than those scoring low. The hypothesis was only partially supported by results. The results suggest that foreign students with low English verbal aptitude can succeed in American graduate schools. The limitations of GPA as a criterion of graduate school success for foreign students is discussed. (Author)

EDO 58304

RB-71-50

RESEARCH BULLETIN

TEST OF ENGLISH AS A FOREIGN LANGUAGE AS A MODERATOR OF GRADUATE RECORD EXAMINATIONS SCORES IN THE PREDICTION OF FOREIGN STUDENTS' GRADES IN GRADUATE SCHOOL

Amiel T. Sharon

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TM 001 019

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
September 1971

TEST OF ENGLISH AS A FOREIGN LANGUAGE AS A MODERATOR OF GRADUATE RECORD EXAMINATIONS SCORES IN THE PREDICTION OF FOREIGN STUDENTS' GRADES IN GRADUATE SCHOOL

Amiel T. Sharon

Abstract

Scores of foreign graduate students on the Graduate Record Examinations (GRE) Aptitude Tests and the Test of English as a Foreign Language (TOEFL) were combined through multiple and moderated regression to predict grade-point average (GPA). It was hypothesized that TOEFL would moderate the relationship between the GRE scores and GPA. According to this hypothesis, students scoring high on TOEFL would be more predictable by GRE than those scoring low. The hypothesis was only partially supported by the results. The results suggest that foreign students with low English verbal aptitude can succeed in American graduate schools. The limitations of GPA as a criterion of graduate school success for foreign students is discussed.

TEST OF ENGLISH AS A FOREIGN LANGUAGE AS A MODERATOR OF GRADUATE RECORD EXAMINATIONS SCORES IN THE PREDICTION OF FOREIGN STUDENTS' GRADES IN GRADUATE SCHOOL

Amiel T. Sharon

Recent years have seen a significant growth in the number of students from foreign countries enrolling in American graduate schools, and this trend is likely to continue in the future. The annual foreign student census indicates that there were 54,034 foreign graduate students enrolled in American universities in the 1968-69 academic year (Institute of International Education, 1969). The current enrollment is more than twice that of a decade ago.

The admission of foreign students to graduate study in the United States is a complex problem. Unlike their American counterparts, foreign students often lack proficiency in the English language and have different language and cultural backgrounds. Furthermore, undergraduate record, which generally has been found to be the best predictor of graduate school success, is difficult to evaluate for the foreign student. The lack of comparability in the grading systems of universities in different countries makes it impossible to employ the prediction approach used with American students. The appraisal of the foreign candidate's aptitude for graduate study by standardized admissions tests also has pitfalls. Poor performance may be due to factors not directly related to aptitude for graduate study. For example, the nonnative examinee may lack adequate English proficiency to understand the test questions or he may not be familiar with the philosophy or method of American objective tests.

Competence in the English language is one factor which has been assumed to be crucial for the success of the foreign student studying at an American university. It is difficult to imagine how a student can learn in an American graduate school without being able to read, write, and comprehend in the English language. Thus, English proficiency might be thought of as a necessary, although not sufficient, prerequisite for graduate school success. For this reason many graduate schools recommend or require that their foreign applicants take the Test of English as a Foreign Language (TOEFL) in their native country prior to coming to the United States.

TOEFL is designed to help foreign students demonstrate their English language proficiency at the advanced level required for study at American colleges and universities. The test consists of five parts, each part measuring an important language skill: (1) Listening Comprehension, (2) English Structure, (3) Vocabulary, (4) Reading Comprehension, and (5) Writing Ability. A total score is reported in addition to the five part scores.

A requirement of many graduate schools of all their applicants is that competence for advanced study be demonstrated by successful performance on the aptitude test of the Graduate Record Examinations (GRE) (Lannholm, 1967). The two aptitude tests, Verbal (V) and Quantitative (Q), are designed to measure mental capabilities thought to be important in graduate level study. They are not achievement or proficiency tests which require knowledge in any specific subject matter. Instead, they attempt to measure reading comprehension and logical reasoning with both verbal and quantitative material.

There is no information at the present time which would indicate how the combination of scores on TOEFL and GRE are, or should be, used by graduate schools for selecting foreign students. Since TOEFL is labeled a "proficiency" test and the GRE an "aptitude" test, it is logical to assume that the two tests yield different sorts of information about a candidate. For that reason, the combination of the two tests could result in a more accurate prediction of academic achievement than either test alone.

The general purpose of this study was to determine whether TOEFL adds to the predictive validity of the GRE Verbal test. More specifically, it was hypothesized that TOEFL would act as a "moderator" of the relationship between GRE-V and a measure of graduate school performance. The use of moderator variables rests on the assumption that within any given prediction system there are classes or types of individuals who can be predicted with greater accuracy than others. The existence of moderated relationships has been demonstrated in previous studies (e.g., Frederiksen & Gilbert, 1960; Saunders, 1956). It was hypothesized that TOEFL would act as a moderator variable; that is, students scoring high on TOEFL would be more predictable by GRE-V than those scoring low. It would seem reasonable to assume that if an individual does not have adequate English proficiency, a verbal aptitude test could not accurately predict his scholastic achievement. The practical implication is that the GRE-V score of a foreign applicant with a low TOEFL score should be ignored.

A similar argument for the use of TOEFL as a moderator variable has been made by Slocum (1967). He states,

> ...such tests as the SAT and GRE are most useful in the case of foreign applicants when their scores are considered in the light of the TOEFL scores. This enables one to make allowances, if necessary, in objective test scores for a handicap in English. On the other hand, the TOEFL score could point to rejection if it were high (indicating little or no handicap in English) and the objective test scores (not so much verbal aptitude scores as quantitative and achievement scores) were low.

Inconclusive results were obtained in a previous study which tested the hypothesis that TOEFL acts as a moderator variable (Sharon, 1970). TOEFL and the Admissions Test for Graduate Study in Business (ATGSB) were statistically combined to predict foreign student success in business school. It was found that TOEFL added nothing to the predictive validity of the ATGSB when combined with it in a linear or nonlinear fashion. The sample used in that study, however, was a small and a highly selected one; the average TOEFL score of the subjects was at the 81st percentile on foreign student norms. The question whether a verbal aptitude test is valid for low TOEFL scorers was not answered by the study.

In a similar study, Schrader and Pitcher (1970) combined TOEFL and the Law School Admission Test (LSAT) scores to predict law school grades of foreign students. They reported that the joint use of TOEFL and LSAT in linear regression did not improve the accuracy of the prediction obtained over that when either test was used alone. The sample of students employed in this study suffered from the same limitations as the one in the ATGSB study. It was too small and unrepresentative for drawing firm conclusions. The mean TOEFL score of the subjects was more than one standard deviation above the mean of all foreign applicants.

The objective of the present study was to obtain a sufficiently large and representative foreign graduate student sample in order to be able to test the hypotheses that

1. TOEFL will moderate the relationship between the GRE-Verbal aptitude test and graduate school grade-point average (GPA).
2. TOEFL will add significantly to the validity of the GRE-Verbal or Quantitative aptitude tests in predicting graduate school GPA.

The use of the GPA as a criterion of graduate school success has been criticized on a number of counts. Davis (1964) found that it fails to reflect certain desirable student attributes such as ethicality, open-mindedness, and self-insight. The lack of variability of grades at the graduate level was noted in previous studies (see Lannholm, 1967). The relevance of grades as a criterion for academic achievement has been questioned. As the dean of one of the schools invited to participate in this study writes:

> Our experience with the grading of graduate students at many universities leads us to doubt that grades in general correlate well with what might be expected to be indices of success in school. Professors give nearly all A's and B's, with a majority of the former. Major professors given their advisees A's with startling frequency. A student acquires a B grade more often by reason of personal problems of a transient nature than because he is less capable than those who get A's in the same course.

Finally, there is the possibility that GPA is a contaminated criterion for foreign students. Frequently these students are given special remedial work and lighter course loads when their English proficiency is found to be inadequate.

In spite of the criticisms of GPA, it should be pointed out that this measure of success frequently determines whether a student remains in school

and obtains an advanced degree. No other single criterion measure has yet been developed which is clearly more relevant or acceptable than GPA. Furthermore, since GPA is readily available in the files of many graduate schools, little effort and expense are needed to obtain it.

Method

In order to avoid the burdensome testing of students with the two examinations under consideration, an attempt was made to obtain the necessary data from the records of graduate schools which recommend or require their foreign applicants to take both TOEFL and the GRE. One hundred forty schools, each of which enrolled at least 50 foreign students, were contacted by letter. Each school was asked to supply the GRE Aptitude and TOEFL scores and graduate school GPAs for all foreign graduate students presently or previously enrolled back to and including 1967. Information was also sought on the number of semesters on which the GPA was based, whether the student withdrew from the university, and the major field of each student.

Of the 140 schools contacted, 24 schools (see Appendix) provided useable data on a total of 978 foreign students. Seventy other schools responded by either returning unuseable data (e.g., GRE or TOEFL scores missing) or indicating that the requested information was not available or not retrievable from their records.

Results and Discussion

Test and School Performance of Foreign Students

The means and standard deviations of the TOEFL and GRE scores for the study sample and for selected reference samples are indicated in Table 1.

The reference sample for TOEFL consists of 113,975 foreign students seeking admission to institutions of higher education in the United States who took TOEFL from February 1964 through June 1969 (Educational Testing Service, 1970). The reference sample for GRE consists of approximately 539,000 candidates, probably almost all native Americans, who took the Aptitude test from May 1966 through April 1969 (Educational Testing Service, 1969).

The students in the study sample scored, on the average, over one-half of a standard deviation above the mean of all foreign applicants on TOEFL, very likely because they consisted of enrolled students who were selected, at least to some extent, on the basis of their TOEFL scores. The mean GRE scores of the sample indicate that there is a great discrepancy, relative to American students, between their verbal and quantitative abilities as measured by the GRE. As a group, the subjects are more than one standard deviation below the mean on GRE-V but more than one-half of one standard deviation above the mean on GRE-Q. The superior quantitative scores of the subjects appear to be related to the fact that half of them were majoring in subjects requiring extensive use of this ability. The mean GRE-Q score of those majoring in engineering, technology, and mathematics was 670 as compared to the mean of 547 of all the other majors. The discrepancy between the verbal and quantitative scores of foreign students enrolled in American schools was also noted in previous studies (Harvey & Lannholm, 1961; Sharon, 1970).

Table 1

Means and Standard Deviations for Study Sample and

Reference Applicant Samples on TOEFL and GRE

Test	Study Sample		Foreign Applicants	
	Mean	Standard Deviation	Mean	Standard Deviation
TOEFL - Part 1	55	10	50	10
TOEFL - Part 2	53	7	49	8
TOEFL - Part 3	54	9	48	11
TOEFL - Part 4	54	7	48	8
TOEFL - Part 5	53	7	48	8
TOEFL - Total	537	65	487	78
	Study Sample		Native Applicants	
Test	Mean	Standard Deviation	Mean	Standard Deviation
GRE - Verbal	348	96	516	129
GRE - Quantitative	609	128	524	138

The intercorrelations between the GRE and TOEFL scores are shown in Table 2. There is a relatively high correlation (.70) between GRE-V and TOEFL-Total indicating that the two tests are, to a large degree, measuring the same ability or proficiency. However, since the reliability of GRE-V is .93 (Educational Testing Service, 1969) and that of TOEFL-Total is .97 (Educational Testing Service, 1970), the tests can hardly be taken to be parallel measures of the same linguistic skills.

Table 2

Intercorrelations of GRE and TOEFL Scores

	GRE	
	V	Q
TOEFL - Part 1	.58	.14
TOEFL - Part 2	.45	.25
TOEFL - Part 3	.58	.14
TOEFL - Part 4	.64	.37
TOEFL - Part 5	.48	.25
TOEFL - Total	.70	.26

The scholastic performance of the subjects, as represented by GPA, appears to have been excellent. Only three out of 978 students were reported to have withdrawn from graduate school. Since the test scores of these three students were not included in the analyses, generalizations of the results of this study are limited to those foreign students who remain in graduate school at least one semester. The GPAs of all but three schools were on a five-point scale where A=4 and F=0. In order to make the grading units comparable across all schools, the grades of these three schools were converted to the five-point scale.

The mean GPA of all students was 3.39 with a SD of .45. About 87 per cent of the students earned a B average or higher. The GPAs of the subjects were not all based on the same number of semesters of study. Twenty-nine per cent of the GPAs were based on one semester, 21 per cent on two semesters, 19 per cent on three semesters, and the remaining 31 per cent on four or more semesters.

Central Prediction

The use of grades as the criterion of graduate school success made it necessary to consider the problem of different grading standards at the participating schools. Since a grade of A in one school might be equivalent to a grade of B in another school, serious error may be introduced into any prediction system that did not adjust for these differences. The simplest way of solving this problem might be to establish a separate regression system for each school. This solution, however, requires that a large enough sample of students be tested at each school so that stable regression coefficients would be established. Tucker (1963) developed a central prediction system useful for pooling data across a number of schools in order to increase the sample size for meaningful regression analysis. The central prediction system can be used to compute one set of regression weights which apply to all schools. The differential grading problem is solved by introducing additive and multiplicative constants for adjusting the predicted grades in each school. These two constants are determined in part by the variability and average level of a particular school's GPA distribution. The central regression weights are determined in conjunction with the school constants such that a least squares error function is minimized. The regression weights and validity for any particular school are determined by information unique to that school and also determined by information derived from all the other schools within the system. The information in the system but outside of the particular school supplements and stabilizes the validity within the school so that minimum shrinkage results under cross-validation.

The initial analysis consisted of combining GRE-V or Q with TOEFL-Total in a linear multiple regression through the central prediction system. Since there was a possibility that different abilities would be required for success in different fields, the central prediction analyses were conducted by major field in those fields which had a sufficient number of subjects. Table 3 indicates the number and percentage of subjects in each major field and their test scores and GPAs. The catch-all category "other" consists of all students not majoring in engineering, technology, mathematics, or natural sciences.

Table 3

Test and School Performance Means of Subjects by Major Field

			TOEFL						GRE		
Major Field	N	%	1	2	3	4	5	Total	V	Q	GPA
Engineering, Technology, & Mathematics	492	50	55	54	55	55	53	544	360	670	3.47
Natural Sciences	176	18	52	52	53	52	52	522	320	610	3.32
Other	307	32	55	53	54	52	52	534	343	511	3.31
All Subjects	975	100	55	53	54	54	53	537	348	609	3.39

Table 4 indicates the average validities (weighted by the number of cases at each school) of the predictors and certain predictor composites by major field. It can be seen in Table 4 that the best single overall

predictor is GRE-Q with a validity coefficient of .32 for all subjects. Only in the major field category of "other" is GRE-Q less valid than TOEFL. TOEFL, however, with a validity of .39 in the "other" category is not significantly different from GRE-V with a validity of .35 ($t = 1.76$, $p > .05$). It is also apparent from Table 4 that the linear combinations of GRE-V or Q with TOEFL do not result in significantly higher validities over those obtained when a single best predictor is used alone.

Table 4

Validities of Predictors and Predictor

Composites by Major Field

Major Field	GRE-V	GRE-Q	TOEFL	TOEFL & GRE-V	TOEFL & GRE-Q
Engineering, Technology, and Mathematics	.22	.39	.21	.23	.39
Natural Sciences	.41	.59	.39	.42	.61
Other	.35	.28	.39	.39	.39
All Subjects	.24	.32	.26	.27	.34

The size and the pattern of the GRE validities generally agree with those found in a previous validity study with foreign students (Harvey & Pitcher, 1963). In that study, average correlations between GRE-V and Q and GPA for 637 foreign graduate students were found to be .22 and .28 respectively. In this study the corresponding correlations are .24 and .32. One possible explanation for the higher validity of the Quantitative test is the relatively greater restriction in the range of GRE-V. Another

explanation is that the result is simply a function of the major fields in which most foreign students are enrolled. Consistent with the latter interpretation is the finding that TOEFL and GRE-V have higher validities than GRE-Q for the "other" major field.

Several reasons may be offered for the finding that both GRE-V and TOEFL in general correlate lower with grades than does GRE-Q. First, the sample consists mostly of students majoring in fields in which quantitative ability is likely to play a more important role than verbal ability. Second, it is quite likely that the GPA is an inappropriate criterion for foreign students. For example, it is generally known that foreign students with inadequate English proficiency are frequently given special remedial English instruction and lighter course loads. One might also suspect that allowances are often made in the grades themselves for inability of foreign students to deal comfortably with the American language, concepts, and culture. All this adds up to a bad case of criterion contamination which tends to reduce the correlation between the test scores and grades. However, since criteria of graduate school success other than grades are difficult to find, further study of the effect of criterion contamination is needed.

Further analysis of the data was made to determine whether TOEFL moderates the relationship between GRE and GPA. Three equal sized English proficiency groups (low, middle, and high) based on total TOEFL score were formed within the major fields of engineering, technology, and mathematics and "other" (the natural sciences group did not consist of sufficient cases for analysis). The mean weighted validities of the GRE aptitude tests within the subgroups are presented in Table 5. The validities for some of

the subgroups are substantially higher than the corresponding validities for the total group. In the major field of engineering, technology and mathematics the validity of GRE-V is raised from .22 to .35 in the low proficiency group and to .36 in the middle proficiency group. In the same major field the validity of GRE-Q is raised from .39 to .56 in the low proficiency group. In the "other" major field the validity of GRE-V is raised from .35 to .44 in the middle proficiency group and that of GRE-Q increased from .28 to .35 in the middle proficiency group and to .37 in the high proficiency group. Although these results suggest that TOEFL may be a moderator of the GRE in the prediction of graduate GPA, the hypothesis that high TOEFL scorers are more predictable by the GRE is only partially supported by the results for the "other" group. In the engineering, technology, and

Table 5

GRE Aptitude Test Validities for Low, Middle,

and High English Proficiency Groups

Major Field	English Proficiency					
	Low		Middle		High	
	GRE-V	GRE-Q	GRE-V	GRE-Q	GRE-V	GRE-Q
Engineering, Technology, & Mathematics	.35	.56	.36	.42	.21	.43
Other	.30	.25	.44	.35	.38	.37

mathematics group the opposite of what was predicted resulted. The low proficiency group is apparently more predictable by either aptitude test than the high proficiency group.

Perhaps the most noteworthy finding of this study is that, in general, foreign students appear to succeed in American graduate schools in spite of scoring more than one standard deviation below the mean of American students on GRE-V. This finding suggests that the scores of foreign applicants on this test should be interpreted cautiously when evaluating their aptitude for graduate study.

Conclusion

The results of this study indicate that an English proficiency test such as TOEFL may raise the validity of the GRE aptitude tests in predicting foreign students' graduate school GPA. However, the hypothesis that high TOEFL scorers would be more predictable by GRE than those less proficient in English is only partially supported by the results. Furthermore, it appears that foreign students with low English verbal aptitude can succeed in American graduate schools.

It should be noted that both the GRE and TOEFL may have other purposes besides the prediction of GPA. For example, TOEFL consists of five separate subtests designed to be appropriate in difficulty for foreign candidates. It may, therefore, have diagnostic value in uncovering English language deficiencies of a candidate. The GRE provides verbal and quantitative scores which might be useful for guidance to a particular major field. This study investigated the joint usefulness of the two tests only for predicting GPA which may be a questionable criterion of graduate school success for foreign students.

References

Davis, J. A. Faculty perceptions of students. II. Faculty definition of desirable student traits. Research Bulletin 64-11. Princeton, N. J.: Educational Testing Service, 1964.

Educational Testing Service. Guide to the use of GRE scores in graduate admissions. Princeton, N. J.: ETS, 1969.

Educational Testing Service. Test of English as a Foreign Language: Interpretive information. Princeton, N. J.: ETS, 1970.

Frederiksen, N., & Gilbert, A. C. F. Replication of a study of differential predictability. Educational and Psychological Measurement, 1960, 20, 759-767.

Harvey, P. R., & Lannholm, G. V. The performance of foreign graduate students on the Graduate Record Examination Aptitude Test. Graduate Record Examinations Special Report 61-1. Princeton, N. J.: Educational Testing Service, 1961.

Harvey, P. R., & Pitcher, B. The relationship of Graduate Record Examinations Aptitude Test scores and graduate school performance of foreign students at four American graduate schools. Graduate Record Examinations Special Report 63-1. Princeton, N. J.: Educational Testing Service, 1963.

Institute of International Education. Open Doors 1969. New York: Institute of International Education, 1969.

Lannholm, G. V. Graduate schools requiring or recommending GRE scores for initial admission to graduate study. Graduate Record Examinations Special Report 67-3. Princeton, N. J.: Educational Testing Service, 1967.

Saunders, D. R. Moderator variables in prediction. Educational and Psychological Measurement, 1956, 16, 209-222.

Schrader, W. B., & Pitcher, B. Interpreting performance of foreign law students on the Law School Admission Test and the Test of English as a Foreign Language. Statistical Report 70-25. Princeton, N. J.: Educational Testing Service, 1970.

Sharon, A. T. The effectiveness of TOEFL and ATGSB as combined predictors of foreign student success in graduate business school. Unpublished manuscript, Educational Testing Service, 1970.

Slocum, J. B. Using TOEFL in foreign student admissions. Talk given at the 1967 Annual Meeting of the College Entrance Examination Board.

Tucker, L. R. Formal models for a central prediction system. Psychometric Monograph No. 10. Richmond, Va.: William Byrd Press, 1963.

Appendix

Participating Graduate Schools

Brown University
Fort Hays Kansas State College
Fresno State College
Kent State University
Montana State University
Northern Illinois University
Northwestern University
Notre Dame University
Oklahoma State University
Stanford University
State University of New York at Albany
Texas A&M University
University of California (Los Angeles)
University of Connecticut
University of Florida
University of Kentucky
University of Michigan
University of Nevada
University of North Carolina at Chapel Hill
University of Rhode Island
University of Utah
University of South Carolina
The University of Iowa
The University of Texas at Austin